Curriculum Vitae

NAME: HUA XU

Office Address: 7000 Fannin St, Suite 870, Houston, TX 77030
Office Phone: 713-500-3924
Email: hua.xu@uth.tmc.edu

EDUCATION

Ph.D. in Biomedical Informatics **2008**
Columbia University, New York, NY

M.A. in Biomedical Informatics **2004**
Columbia University, New York, NY

M.S. in Computer Science **2001**
New Jersey Institute of Technology, Newark, NJ

B.S. in Biochemistry **1998**
Nanjing University, Nanjing, P. R. China

ACADEMIC APPOINTMENTS

2015 - present	Professor School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX
2012 – present	Director, Center for Computational Biomedicine Robert H. Graham Professorship in Entrepreneurial Biomedical Informatics and engineering CPRIT Scholar in Cancer Research School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX
2012 – 2015	Associate Professor School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX
2008 – 2012	Assistant Professor Department of Biomedical Informatics, School of Medicine, Vanderbilt University, Nashville, TN

PROFESSIONAL ORGANIZATIONS

2005 – Present	American Medical Information Association (AMIA)
2005 – Present	The Association for Computational Linguistics (ACL)

PROFESSIONAL ACTIVITIES

Intramural Committees:

2014 – present	Interfaculty Council, The University of Texas Health Science Center at Houston
2014 - present	Bylaw Committee, School of Biomedical Informatics, The University of Texas Health Science Center at Houston
2013 - present	Faculty Promotion Committee, School of Biomedical Informatics, The University of Texas Health Science Center at Houston
2012 - present	Faculty Search Committee, School of Biomedical Informatics, The University of Texas Health Science Center at Houston
2012 - present	Executive Committee, Innovation for Cancer Prevention Research (ICPR) Training Program, The University of Texas School of Public Health
2012	BioVU (Vanderbilt's Biobank) Review Committee, Vanderbilt University
2010 - 2012	Academic Program Committee of Biomedical Informatics, Vanderbilt University

Extramural Committees:

Board of Scientific Counselors, Lister Hill Center, National Library of Medicine, NIH

2014 – present	Board member

AMIA Natural Language Processing (NLP) Working Group:

2013 – 2014	Chair
2012	Chair-Elect

Conference Program Committee Chair/Track Chair:

2014	The ACM Eighth International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO'14), in conjunction with the ACM CIKM, Shanghai, China
2014	American Medical Informatics Association (AMIA) NLP Working Group pre-symposium – a doctoral consortium and a workshop on interoperability, Washington DC, US
2014	American Medical Informatics Association (AMIA) Translational Bioinformatics Summit, San Francisco, US, SPC Track-Chair
2013	The ACM Seventh International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO'13), in conjunction with the ACM CIKM, San Francisco, US

2013	American Medical Informatics Association (AMIA) NLP Working Group pre-symposium – a doctoral consortium and a workshop on open source clinical NLP systems, Washington DC, US
2012	International Conference on Intelligent Biology and Medicine (ICIBM), Nashville, Tennessee, US
2012	The ACM Sixth International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO'12), in conjunction with the ACM CIKM, Hawaii, US
2012	American Medical Informatics Association (AMIA) NLP Working Group pre-symposium – a doctoral consortium and a data workshop, Washington DC, US

Conference Program Committee Member:

2015	Biomedical Linked Annotation Hackathon (BLAH2015), Kashiwa-no-ha City, Japan
2015	American Medical Informatics Association (AMIA) Translational Bioinformatics Summit, San Francisco, US
2014	International Conference on Intelligent Biology and Medicine (ICIBM'14), San Antonio, Texas, US
2014	International Workshop on Biomedical and Health Informatics, IEEE BIBM, Belfast, UK
2014	IEEE International Conference on Healthcare Informatics (ICHI), Verona, Italy
2013	International Workshop on Biomedical and Health Informatics, IEEE BIBM, Beijing, China
2013	IEEE International Conference on Healthcare Informatics (ICHI), Philadelphia, PA, USA
2013	International Conference on Intelligent Computing (ICIC'13), Nanning, China
2013	International Conference on Intelligent Biology and Medicine (ICIBM'13), Nashville, Tennessee, US
2012	AMIA Annual Fall Symposium
2012	2nd ACM SIGHIT International Health Informatics Symposium, Miami, Florida, US
2012	IEEE Conference on Healthcare Informatics, Imaging, and System Biology (HISB), San Diego, US
2011	American Medical Informatics Association (AMIA) Translational Bioinformatics Summit, San Francisco, US
2011	IEEE Conference on Healthcare Informatics, Imaging, and System Biology (HISB), San Jose, US
2011	International Conference on Intelligent Computing (ICIC'12), Zhengzhou, China
2011	International Biomedical Informatics Summit at Peking University, Beijing, China
2011	International Workshop on Biomedical and Health Informatics, IEEE BIBM, Atlanta, Georgia, US

Grant Reviewer:

2014	The South African Medical Research Council

2014	Medical Research Council (MRC), UK
2014	Fragile X review, National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH)
2013	Reviewer at Kentucky Science and Engineering Foundation
2013	Ad-hoc reviewer for GNOM-G Review Panel, National Human Genome Research Institute (NHGRI), National Institutes of Health (NIH)
2012	Genomic Medicine Pilot Demonstration Projects, RFA-HG-12-006, National Human Genome Research Institute (NHGRI), National Institutes of Health (NIH)
2012	Epidemiology and genetics of chronic diseases, PA-11-260, National Cancer Institute (NCI), National Institutes of Health (NIH)
2012	Mail reviewer, NIH Director's Early Independence Award (DP5), RFA-RM-11-007, National Institutes of Health (NIH)
2011	PFINDR: Phenotype Finder IN Data Resources: A Tool to Support Cross-study Data Discovery Among NHLBI Genomic Studies (UH2/UH3), RFA-HL-11-020, National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH)
2011	Conference Grant Applications (R13), PA-10-07, National Institute on Drug Abuse (NIDA), National Institutes of Health (NIH)

Editorial Activities:

Editorial Board Member:

2013 – present	Journal of American Medical Informatics Association (JAMIA)
2012 - present	Journal of Biomedical Informatics (JBI)
2010 - present	International Journal of Data Mining and Bioinformatics (IJDMB)

Guest Editor:

2012	BMC Systems Biology Supplement issue "Advances in Intelligent Biology and Medicine: Selected papers from 2012 ICIBM Conference". 6 (Suppl S2)
2012	BMC Genomics Supplement issue "Advances in Intelligent Biology and Medicine: Selected papers from 2012 ICIBM Conference". 13 (Suppl S7)
2012	International Journal of Computational Biology and Drug Design (IJCBDD) special issue "Advances in Intelligent Biology and Medicine: Selected papers from 2012 ICIBM Conference". Volume 5, issue 4

Journal Reviewer:

Bioinformatics, Journal of American Medical Informatics Association, BMC Bioinformatics, BMC Medical Informatics and Decision Making, Journal of Biomedical Informatics, Journal of Biomedical Discovery and Collaboration, Methods of Information in Medicine, Transactions on Information Technology in BioMedicine, Journal of Medical Internet Research, PLos ONE

Conference Reviewer:

2009 – 2014	AMIA TBI & CRI Summit
2007 - 2014	AMIA Fall Symposium
2012	2nd ACM SIGHIT International Health Informatics Symposium, Miami, Florida, US

2011	International Workshop on Biomedical and Health Informatics, IEEE BIBM, Atlanta, Georgia, US
2011	IEEE Conference on Healthcare Informatics, Imaging, and System Biology (HISB), San Jose, US
2011	International Conference on Intelligent Computing, Zhengzhou, China
2007	ACL BioNLP Workshop
2007	Symposium on Languages in Biology and Medicine (LBM)

Honors/Awards:

2014	Elected **Fellow American College of Medical Informatics (ACMI)**
2014	Ranked #1, SemEval-2014 Task 7 – Analysis of Clinical Text
2013	**IBM Faculty Award**
2013	Top-scoring predictions for the NIEHS-NCATS-UNC DREAM Toxicogenetics Challenge leaderboard
2013	Top ranked: Task 1a – 1st, Task 1b – 3rd, and Task 2 – 1st, at 2013 ShARe/CLEF eHealth Shared Tasks in Clinical NLP
2012	Frist in Temporal Relation Extraction tasks, 2012 i2b2 Clinical NLP challenge
2012	Best Paper of the Year by the International Medical Informatics Association, PMC3237759, 2012 Yearbook in Medical Informatics
2012	Finalist of best paper award, AMIA Translational Bioinformatics Summit
2011	Finalist of distinguished paper award, AMIA Fall Symposium
2011	Best Paper of the Year by the International Medical Informatics Association, PMC2815478, 2011 Yearbook in Medical Informatics
2010	2nd best system in Concept Extraction task, 2010 i2b2 Clinical NLP challenge
2009	**Homer R. Warner Award, AMIA Fall Symposium**
2009	Distinguished paper award (co-author), AMIA Fall Symposium
2009	2nd best system, 2009 i2b2 Clinical NLP challenge
2007-2008	National Library of Medicine Training Fellowship

Invited Talks/Presentations:

Invited Talks:

2014	JAMIA Journal Club, "*Validating drug repurposing signals using EHRs and informatics: a use case of metformin associated with reduced cancer mortality*"
2014	Wayne State University, Detroit, Department of Computer Science, "*Natural language processing in the medical domain: Methods and applications*"
2014	Suzhou University, China, Center for Systems Biology, "*An introduction to natural language processing in the medical domain*"
2014	Cancer Hospital, Shanghai, China, "*Cancer research using electronic health records.*"

2014	Harbin Institute of Technology, China, Department of Computer Science (Shenzhen), “*Methods and applications of natural language processing in the medical domain*”
2014	International Conference on Frontiers in Chronic Disease Research and Prevention, Shanghai, “*Electronic health records and informatics for epidemiological studies on chronic diseases*”
2013	The University of Texas Academia-Industry Partnership FreshAir Roundtable, “*The use of electronic health records for repurposing drugs for cancers*”
2013	The eMERGE consortium, “*MedEx-UIMA – an open source system for medication information extraction*”.
2013	University of San Diego, Division of Biomedical Informatics, iDASH center, “*Medication information extraction from EHR – methods and applications.*”
2013	The Methodist Hospital Research Institute, Department of Systems Medicine and Bioengineering, “*Natural language processing: methods and applications in clinical research*”
2012	University of Wisconsin - Milwaukee, Center for Biomedical Data and Language Processing, “*Studying drug outcomes using electronic medical records and informatics approaches*”
2012	Pharmacogenomics Research Network (PGRN), Research-In-Progress Seminar, “*Pharmacogenomic discovery and replication using EHR, biobank, and informatics - the PGPop resource”*
2012	University of Michigan, Ann Anbor, Department of Computational Medicine and Bioinformatics, “*Studying drug outcomes using electronic medical records and informatics approaches*”
2012	The Ohio State University, Columbus, Department of Biomedical Informatics, “*Studying drug outcomes using electronic medical records, biobanks, and informatics*”
2012	Indiana University, School of Medicine, Center for Computational Biology and Bioinformatics, “*Pharmacogenomic studies using EHR, Biobank, and Informatics*”
2011	University of Alabama at Birmingham, Section on Statistical Genetics, “*Pharmacogenetic studies using electronic medical records, biobank, and informatics*”
2011	International Biomedical Informatics Summit at Peking University, Beijing, China, “*Linking EHR to DNA biobank for genomic research: an informatics perspective*.”
2011	University of Texas Health Center at Houston, School of Biomedical Informatics, “*Pharmacogenomic discovery using electronic medical records, biobank, and informatics*”
2011	Vanderbilt University, Department of Medicine, “Dinner and Data” Seminars, “*Informatics approaches for EHR-based pharmacogenomics studies”*
2010	Peking University, China, Center of Medical Informatics, *“Research and Education in Medical Informatics at Vanderbilt University*”
2010	Shanghai Center for Bioinformation Technology, China, *“Research and Education in Medical Informatics at Vanderbilt University*”
2010	Medical University of South Carolina, Department of Biochemistry, *“Facilitating EMR-based Pharmacogenetic Studies using Informatics Methods*”

2010	George Town University, CTSA informatics division, “*Electronic Medical Records, Informatics, and Clinical Research*”
2009	eMERGE network meeting, “*Medication extraction from clinical text*”
2008	University of Kansas, Department of Electrical Engineering and Computer Science, “*Natural Language Processing in the Biomedical Domain*”
2008	Arizona State University, Department of Biomedical Informatics, *“A Framework for Handling Biomedical Abbreviations*”
2008	Ohio State University, Department of Biomedical Informatics, *“A Framework for Handling Biomedical Abbreviations”*

Presentations:

2014	AMIA Fall Symposium, “*Mining electronic health records to detect drug repurposing singals for cancers”*
2014	AMIA Clinical Research Summit, “*Extracting and standardizing medication information in clinical text - the MedEx-UIMA system”*
2013	Medinfo. “*Analyzing differences between Chinese and English clinical text: a cross-institution comparison of discharge summaries in two languages”*
2013	ACM Sixth International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO), “Clinical Entity Recognition using Structural Support Vector Machines with Rich Features”
2013	AMIA Annu Symp Proc. “*Combining corpus-derived sense profiles with estimated frequency information to disambiguate clinical abbreviations*”
2012	Pac Symp Biocomput, *“Ranking gene-drug relationships in biomedical literature using latent dirichlet allocation”*
2011	AMIA Translational Bioinformatics Summit, “*Facilitating Electronic Health Records based Pharmacogenetic Studies using Natural Language Processing: a case study of warfarin.”*
2011	AMIA Clinical Research Informatics, “*A Framework to Determine Patient Drug Exposure from EMR: an Application to warfarin.”*
2011	International Workshop on Biomedical and Health Informatics, IEEE Conference of Bioinformatics and Biomedicine (BIBM), “*An Initial Study of Full Parsing of Clinical Text using the Stanford Parser.*”
2010	Workshop of i2b2 NLP Challenge, “*Hybrid Approaches to Concept Extraction and Assertion Classification – Vanderbilt’s systems for 2010 i2b2 NLP challenge*”
2009	Workshop of i2b2 NLP Challenge, “*Vanderbilt’s System for Medication Extraction*”.
2009	AMIA Fall Symposium, “*MedEx – A Medication Information Extraction System for Clinical Narratives*”
2008	AMIA Fall Symposium, “*Methods for Building Sense Inventories of Abbreviations in Clinical Notes*”
2007	AMIA Fall Symposium, “*A Study of Abbreviations in Clinical Notes*”

Tutorials:

2013	MedInfo 2013, “Clinical Natural Language Processing”
2013	AMIA Annual Symposium, “*Using Electronic Health Records to facilitate clinical, genomic, and pharmacogenomic research: Experiences from the eMERGE and PGRN Networks”*

2012	AMIA Translational Bioinformatics Summit, "*Reusing EHRs for Clinical, Genomic, and Pharmacogenomic Discovery at Vanderbilt and within the eMerge Network*"
2011	AMIA Now, "*NLP 101: Making sense of EMR text*"

Panels:

2014	iDASH annual conference, San Diego, "*Making biomedical data discoverable*"
2013	AMIA Fall Symposium, "*Informatics to enable routine* personalized cancer therapy"
2012	Academy Health Annual Research Meeting, "*Squeezing Blood from a Stone: Principles and Practice of Using Narrative Data in Health Services Research*"
2012	NIH workshop on "Natural Language Processing: State of the Art, Future Directions and Applications for Enhancing Clinical Decision-Making***",*** Session Chair of "*Linguistics-based methods*" panel.
2012	AMIA Clinical Research Informatics Summit, "*Natural Language Processing for Clinical and Translational Research"*
2012	Academy Health Annual Research Meeting, Health IT Interest Group, "*HIT Quality of Care and Reporting Panel"*
2011	AMIA Clinical Research Informatics Summit, "*Cross-Institutional Systems to Support Phenotyping in Biomedical Research – Experience from the emerge network*"

TEACHING ACTIVITIES

Courses:

2013 - 2014	HI6330 Biomedical Natural Language Processing: Methods and Applications, instructor
2010 - 2012	BMIF 300 Introduction to Biomedical Informatics, lecturer
2009 - 2012	BMIF-315 Methodological Foundations of Biomedical Informatics, lecturer
2009	BMIF-330 Machine Learning for Biomedicine, co-instructor

Research Advisor:

Postdoctoral Fellow:

2014 – present	Jun Xu, Ph.D.
2013 – present	Ergin Soysal, M.D., Ph.D.
2013 – present	Yaoyun Zhang, Ph.D.
2013 – present	Sungrim Moon, Ph.D.
2013 – 2014	Guocai Chen, Ph.D.
2011 - 2013	Buzhou Tang, Ph.D.
2010 - 2014	Yonghui Wu, Ph.D.
2009 - 2012	Mei Liu, Ph.D.
2010 - 2011	Samir AbdelRahman, Ph.D.

2009 - 2010	Son Doan, Ph.D.

PhD Student:

2014 – present	Qiang Wei, PhD student
2013 – present	Jun Li, PhD Candidate
2012 - present,	Min Jiang, PhD Candidate,
2011 - 2015	Yukun Chen, PhD
2012 – 2014	Jianbo Lei, PhD (Dissertation title: Named entity recognition in Chinese clinical text)

MS Student:

2014 – present	Michelle Scerbo, MS Candidate
2014 – present	Ginger Schirmer, MS Candidate
2014 – present	James Schlebus, MS Candidate
2014 – present	Weiming Shi, MS Candidate
2013 – 2014	Anupama E. Gururaj, MS Candidate,
2013 – 2014	Ruiling Liu, MS Candidate
2013 – 2014	Heidi Jones, MS Candidate

Summer Intern:

2015 Summer	Manu Prakasam, High school student
2014 Summer	Kevin Zhu, BS
2013 Summer	Bonnie Nortz, Undergraduate student
2010 Summer	Oluwatoyin Arije, MPharm Candidate
2010 Summer	Kamilah McKinnon, MPharm Candidate

Visiting Student/Scholar:

2015-2016	Hui Chen, PhD
2014-2015	Yubo Zhou, PhD
2011-2012	Hongxin Cao, MS
2009-2010	Yanxin Lu, PhD Candidate

Thesis/Dissertation Committee Member for:

2014 - present	Qiang Wei, PhD student, University of Texas Health Science Center at Houston
2014 - present	Muhammad Amith, PhD student, University of Texas Health Science Center at Houston
2014 - present	Jun Li, PhD student, University of Texas Health Science Center at Houston
2013 - present	Safa Fathiamini, PhD student, University of Texas Health Science Center at Houston
2013 - present	Melissa Resnick, PhD candidate, University of Texas Health Science Center at Houston
2013 - present	Frank Manion, PhD student, University of Texas Health Science Center at Houston
2013 - 2014	Ning Shang, PhD, University of Texas Health Science Center at Houston

2012 - 2013	Sina Madani, PhD, University of Texas Health Science Center at Houston
2012 - 2013	Ravi V. Atreya, MS, Vanderbilt University
2011 - 2014	Robert Carroll, PhD candidate, Vanderbilt University
2009 - 2011	Firas Wehbe, PhD, Vanderbilt University
2010	Jerome Jourquin, MS, Vanderbilt University
2009	Xinxin Zhu, PhD, Columbia University

RESEARCH PROGRAM

Current Grants:

<u>Interactive machine learning methods for clinical natural language processing</u>
NLM 2R01LM010681-05 (PI – Hua Xu)
09/29/2014 – 09/28/2018
Total direct costs: $1,396,822
Role: PI

In this study, we propose to investigate interactive machine learning (IML) methods to address the challenges in clinical NLP about building annotated corpora and combining domain knowledge and statistical learning methods. We will conduct IML studies to three NLP related tasks including word sense disambiguation, named entity recognition, and clinical phenotyping.

<u>Informatics Tools for Pharmacogenomic Discovery using Practice-based Data</u>
NIGMS 1 R01 GM103859-01 (PI: Denny, Pathak, and Xu)
09/18/2014 – 5/31/2018
Total direct budget: $1,588,692
Role: PI
In this study, we will develop natural language processing (NLP) and ontology tools to facilitate pharmacogenomics studies using EHR-linked biobanks.

<u>BioCADDIE: Biomedical and healthCAre Data Discovery and Indexing Engine center</u>
NIH 1U24HL126126-01 (PI – Lucila Ohno-Machado)
09/29/2014 – 09/38/2017
Total direct costs: $6,000,000 (UTHealth $1,025,558)
Role: Subcontract PI, co-investigator

BioCADDIE is a consortium of data producers, curators, publishers, and consumers who will work together to develop practical, sustainable solutions to the problem of biomedical and healthcare data discovery. This project is to develop an NIH BD2K Data Discovery Index Coordination Consortium.

<u>Natural language processing for clinical and translational research</u>
NIGMS 1R01GM102282 (MPI – Hongfang Liu, Serguel Pakhomov, and Hua Xu)
04/01/2013 – 03/31/2017

Total direct costs: $1,345,476 (UTHealth $524,560)
Role: PI

Natural Language Processing (NLP) has played a critical role in enabling secondary use of EHRs data for clinical and translational research. This study will facilitate efficient adoption of NLP to large-scale clinical studies by developing an open source framework that will enable semantic level interoperability among different NLP systems through defined standards and provide user-centric tools for clinical information extraction.

CPRIT Rising Stars Award - Repurposing Existing Drugs for Cancer Treatment using Electronic Health Records
Cancer Prevention & Research Institute of Texas, R1307 (PI – Hua Xu)
03/01/2013 – 02/28/2018
Total direct costs: $2,806,000
Role: PI

This study is to develop novel informatics approaches to facilitate large-scale drug-repurposing studies for identifying potential cancer therapeutic agents by using Electronic Health Records (EHRs) data. The hypothesis is that EHRs can be used to detect new indications of existing drugs for cancer therapy in a very efficient way, with the help of advanced informatics methods.

Patient Medical History Representation, Extraction, and Inference from EHR Data
NLM 1R01LM011829 (PI – Cui Tao)
09/01/2014 – 08/31/2018
Total budget: $1,358,868
Role: co-investigator
This proposed project fills in the current gaps among ontologies, Natural Language Processing (NLP), and EHR-based clinical research for temporal data representation, normalization, extractions, and reasoning. We propose to develop novel approaches for automatic temporal data representation, normalization and reasoning for large, diverse, and heterogeneous EHR data and prepare the integrated data for further analysis.

Learning from patient safety events: a case-based toolkit
AHRQ 1 R01 (PI: Yang Gong)
09/01/2014 – 8/31/2019
Total direct budget: $ 1,246,715
Role: co-investigator
Timely reporting and effective learning from medical incidents is considered an effective way in developing strategies for reducing medical errors. Utilizing an innovative user-centered, learning-supportive, and ontological approach combining with case-based reasoning and natural language processing techniques, we propose to develop a knowledgebase and learning toolkit that can systematically collect and analyze incident reports, linking historical reports with WebM&M, the highest quality of voluntary reports and expert reviews on patient safety. We envision that the innovative approach will facilitate timely, quality reporting and learning from the incidents and ultimately cultivating a just and learning culture of patient safety.

Using Biomedical Knowledge to Identify Plausible Signals for Pharmacovigilance
NLM 1R01LM011563 (PI - Trevor Cohen)
09/01/2013-08/31/2016
Total budget: $611,718
Role: co-investigator

This project will develop novel informatics methods to mine electronic health records and biomedical literature to automatically identify adverse drug events, thus to provide a generalizable approach that can be used to apply knowledge derived from the biomedical literature to interpret clinical data.

Informatics to support routine personalized cancer therapy
NCI 1U01CA180964-01(PI - Elmer Bernstam)
09/01/2013 – 8/31/2016
Total direct budget: $622,500
Role: co-investigator

This project aims to develop an informatics framework to support personalized cancer therapy, which includes tools to 1) implement a bioinformatics pipeline for processing molecular data into actionable profiles, 2) create and maintain a database of therapeutic implications of common genomic aberrations using automated processing of publically-available sources, and 3) develop tools to summarize clinically-relevant genomic alterations encountered in a patient's tumor and genotype-relevant therapies.

Pharmacovigilance Methods: Leveraging Heterogeneous Adverse Drug Reaction Data
NLM R01 LM010016-05 (PI – Carol Friedman)
7/01/2013 – 6/30/2017
Total direct costs: $1,158,496 (UTHealth subcontract $138,000)
Role: Subcontract PI, co-investigator

This project will integrate heterogeneous sources including drug chemical structure, literature, and electronic health records (EHRs) to detect new adverse drug reactions.

Completed Grants:

An In-silico Method for Epidemiological Studies Using Electronic Medical Records
NCI R01CA141307 (PI – Hua Xu)
09/03/2009 – 07/31/2014
Total direct costs: $762,738
Role: PI

The specific aim of this study is to develop an automated informatics approach to extract both fine-grained cancer findings and general clinical information from electronic medical records and use them to conduct cancer related epidemiological studies. It is an EUREKA award (Exceptional, Unconventional Research Enabling Knowledge Acceleration) funded by NCI.

Real-time Disambiguation of Abbreviations in Clinical Notes
NLM R01LM010681 (PI – Hua Xu)

05/31/2010 – 5/30/2014
Total direct costs: $730,000
Role: PI

This project will develop, evaluate, and share a systematic approach to Clinical Abbreviation Recognition and Disambiguation (CARD), and in doing so substantially aims to improve existing NLP systems and computer-based documentation system by reducing ambiguities in electronic records in real-time.

An Informatics-based Approach to Pharmacogenetic Studies of Warfarin
NIH UL1 RR024975-KL2 Scholar Award (PI – Hua Xu)
07/01/2009 – 06/30/2010
Total direct costs: $130,000
Role: PI (70%)

This project is to develop informatics approaches to extract phenotypic data for pharmacogenomics research from EHRs, using natural language processing and machine learning technologies. It was a three-year K Award, but terminated early due to the R01 award.

Bridging Genomics and Medicine by Ontology Fingerprints
NLM 1R56LM010680 – 01A1 (PI – Jim Zheng)
09/01/2012 – 08/31/2014
Total direct costs: $202,898
Role: Co-investigator

This study is to develop an ontology fingerprint - a set of ontology terms overrepresented in the PubMed abstracts linked to a gene or a disease along with the terms' corresponding enrichment p-value, to characterize genes and diseases. The ontology fingerprints are then used to quantify the relationship between a gene and a disease.

National Center for Cognitive Informatics and Decision Making in Healthcare SHARP
ONC 90TR000401 (PI – Jiajie Zhang)
04/10/10-03/31/2014
Total direct costs: $15,000,000
Role: Co-investigator

SHARPc project 3 is to develop informatics methods to automatically create clinical summaries of patients.

Autonomic Cardiovascular Regulation
NHLBI P01HL056693 (PI - David Robertson)
05/01/2012 – 4/30/2017, Participated between 2012-2015
Total direct costs: $5,907,583
Role: Co-investigator

This is a Program Project that is tightly coordinated testing of hypotheses addressing autonomic modulation of cardiovascular function in health and disease. We will use the EMR-linked DNA biobank to study associations between *ADRA2A* genetic variation and increased risk of stress-induced hyperglycemia in patients with myocardial infarction, as well as increased risk of gestational diabetes.

MOMENT (Monitoring for Outpatient Medication Effects and New Toxicities) in TIME
NLM R01 LM007995 (PI - Randy Miller)
02/01/2004 – 06/14/2010
Total direct costs: $697,237
Role: Co-Investigator (10%)

The MOMENT project involves developing sophisticated text-mining and data extraction tools to examine adverse drug effects in patients presenting for emergency and hospital care.

VGER – Vanderbilt Genomic Electronic Medical Records (eMERGE-I)
NHGRI U01 HG004603 (PI - Dan Roden)
09/28/2007 – 07/31/2011
Total direct costs: $6,466,499
Role: Co-Investigator (10%)

The VGER project promotes personalized medicine by conducting genome-phenome association studies using EMR data.

VESPA - Vanderbilt Electronic Systems for Pharmacogenomic Assessment
NIH RC2GM092618 (PI - Dan Masys & Dan Roden)
09/30/2009 – 08/31/2011
Total direct costs: $6,406,203
Role: Co-Investigator (20%)

The VESPA project was a two-year GO grant, which aims to conduct pharmacogenomics studies using electronic health records and DNA biobank.

VGER – Vanderbilt Genomic Electronic Records Project (eMERGE-II)
NHGRI U01 HG006378 (PI - Dan Roden)
08/15/2011 – 07/31/2015
Total direct costs: $4,237,100
Role: Co-investigator (5%), Participated between 2011-2012

The VGER-II project promotes personalized medicine by conducting genome-phenome association studies using EMR data and integrating genomic data into clinical settings.

Pharmacogenomics of Arrhythmia Therapy (PAT)
NHLBI U19HL065962 (PI - Dan Roden)
09/01/2010 – 06/30/2015
Total direct costs: $13,986,828
Role: Co-investigator (30%), Participated between 2010-2012

PAT is one of the sites of Pharmacogenomics Research Network (PGRN). The aim of this project is to understand genetic associations for drugs in arrhythmia therapy. One of the particular projects is to investigate genetic variations associated with warfarin-caused bleedings by using electronic medical records (EMRs) and DNA biobank. We also support PGPop, a network resource for using EMRs and biobanks for pharmacogenomic research.

Evidence-based diagnostic tools for translational and clinical research (eTfor2)
NLM R01LM010828 (PI – Randy Miller)
9/30/2010-9/29/2013
Total direct costs: $730,000
Role: Co-investigator (10%), Participated between 2010-2012

The eTfor2 project develops and evaluates open-source programs and knowledge representations to better characterize patients for translational and clinical research studies.

From GWAS to PheWAS: Scanning the EMR Phenome for Gene-Disease Associations
NLM R01LM010685 (PI - Josh Denny)
09/01/2011 – 8/31/2014
Total direct costs: $652,500
Role: Co-investigator (5%), Participated between 2011-2012

The PheWAS project develops novel NLP and data analysis methods for gene-disease associations based on EMRs. For a given SNP, it scans hundreds of diseases to quickly identify possible associations.

PUBLICATIONS

Peer Reviewed Articles - Journal:

1. **Xu H**, Aldrich MC, Chen Q, Liu H, Peterson NB, Dai Q, Levy M, Shah A, Han X, Ruan X, Jiang M, Li Y, Julien JS, Warner J, Friedman C, Roden DM, Denny JC. Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. *J Am Med Inform Assoc.* 2015, 22(1): 179-91. [PMID: 25053577]
2. Kawai VK, Cunningham A, Vear SI, Van Driest SL, Oginni A, **Xu H**, Jiang M, Li C, Denny JC, Shaffer C, Bowton E, Gage BF, Ray WA, Roden DM, Stein CM. Genotype and risk of major bleeding after the warfarin initiation phase. *Pharmacogenomics*, 2014. Accepted.
3. Shang N, **Xu H**, Rindflesch TC, Cohen T. Identifying plausible adverse drug reactions using knowledge extracted from the literature. *J Biomed Inform.* 2014. Accepted.
4. Huang LC, Soysal E, Zheng W, Zhao Z, **Xu H***, Sun J*. A weighted and integrated drug-target interactome: drug repurposing for schizophrenia as a use case. *BMC Genomics,* 2014. Accepted. (* co-corresponding author)
5. Tang B, Feng, Y, Wang X, Wu Y, Zhang Y, Jiang M, Wang J, **Xu H**. A comparison of conditional random fields and structured support vector machines for chemical entity recognition in biomedical literature. *Journal of Cheminformatics*. 2014 7(Suppl 1):S8 [PMCID: PMC4331698]

6. Zhu K, Liu Q, Zhou Y, Tao C, Zhao Z, **Xu H***, Sun J*. Oncogenes and tumor suppressor genes: comparative genomics and network perspectives. *BMC Genomics,* 2014. Accepted. (* co-corresponding author)
7. Chen Y, Sun, J, Huang LC, **Xu H***, Zhao Z*. Classification of primary cancer sites using machine learning and somatic mutations. *BioMed Research International,* 2014. Accepted. (* co-corresponding author)
8. Niadoo D, Wu AC, Brilliant MH, Denny JC, Ingram C, Kitchner TE, Linneman JG, McGeachie MJ, Roden DM, Shaffer CM, Shah A, Weeke P, Weiss ST, **Xu H**, Medina MW. A polymorphism in HLA-G modifiers statin benefit in asthma. *The Pharmacogenomics Journal.* 2014. Accepted. [PMCID pending]
9. Gobbel GT, Garvin J, Reeves R, Cronin RM, Heavirland J, Williams J, Weaver A, Jayaramaraja S, Giuse D, Speroff T, Brown SH, **Xu H**, Matheny ME. Assisted annotation of medical free text using RapTAT. *J Am Med Inform Assoc.* 2014, 21(5): 833-41. [PMCID PMC4147611]
10. Sun J, Huang L, **Xu H**, Zhao Z. Network-assisted prediction of potential drugs for *addiction. BioMed Research International. 2014, doi:10.1155/2014/258784.* [PMCID pending]
11. Tang B, Cao H, Wang X, Chen Q, **Xu H**. Evaluating word representation features in biomedical named entity recognition tasks. *BioMed Research International*. 2014, doi:10.1155/2014/240403. [PMCID PMC5932722]
12. Liu M, Cai R, Hu Y, Matheny ME, Sun J, Hu J, **Xu H**. Determining molecular predictors of adverse drug reactions with causality analysis based on structure learning. *J Am Med Inform Assoc.* 2014, 21(2):245-51. [PMCID: PMC3932464]
13. Doan S, Lin KW, Conway M, Ohno-Machado L, Hsieh A, Feupeo SF, Garland A, Ross MK, Jiang X, Farzaneh S, Walker R, Alipanah N, Zhang J, **Xu H**, Kim HE. PhenDisco: phenotype discovery system for the database of genotypes and phenotypes. *J Am Med Inform Assoc.* 2014, 21(1):31-6. [PMCID: PMC3912702]
14. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E, Basford MA, Carrell DS, Peissig PL, Kho AN, Pacheco JA, Rasmussen LV, Crosslin DR, Crane PK, Pathak J, Bielinski SJ, Pendergrass SA, **Xu H**, Hindorff LA, Li R, Manolio TA, Chute CG, Chisholm RL, Larson EB, Jarvik GP, Brilliant MH, McCarty CA, Kullo IJ, Haines JL, Crawford DC, Masys DR, Roden DM. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nature Biotechnology*. 2013 31(12):1102-10. [PMCID PMC3969265]
15. Zheng H, Wang H, **Xu H**, Wu Y, Zhao Z, Azuaje F. Linking biochemical pathways and networks to adverse drug reactions. *IEEE Trans Nanobioscience.* 2013: 131-137.
16. Wei WQ, Feng Q, Jiang L, Waitara MS, Iwuchukwu OF, Roden DM, Jiang M, **Xu H**, Krauss RM, Rotter JI, Nickerson DA, Davis RL, Berg RL, Peissig PL, McCarty CA, Wilke RA, Denny JC. Characterization of Statin Dose Response in Electronic Medical Records. *Clin Pharmacol Ther*. 2013, 95(3):331-8. [PMCID PMC3944214]
17. Lei J, Tang B, Lu X, Gao K, Jiang M, **Xu H**. A comprehensive study of named entity recognition in Chinese clinical text. *J Am Med Inform Assoc*. 2014, 21(5):808-14. [PMCID PMC4147609]
18. Van Driest SL, Shah A, Marshall MD, **Xu H**, Smith AH, McGregor TL, Kannankeril PJ. Opioid use after cardiac surgery in children with Down syndrome. *Pediatr Crit Care Med.* 2013, 14(9):862-8. [PMCID: PMC3830692]
19. Fan JW, Yang EW, Jiang M, Prasad R, Loomis RM, Zisook DS, Denny JC, **Xu H***, and Huang Yang*. Syntactic parsing of clinical text: guideline and corpus

development with handling ill-formed sentences. *J Am Med Inform Assoc.* 2013, 20(6):1168-77 [PMCID:PMC3822122]. (* Co-corresponding Authors)

20. Chen Y, Carroll RJ, Shah A, Eyler AE, Denny JC, **Xu H**. Applying active learning to high-throughput phenotyping algorithms for electronic health records data. *J Am Med Inform Assoc.* 2013, 20(e2):e253-9. [PMCID: PMC3861916]
21. Tang B, Wu Y, Jiang M, Chen Y, Denny JC, **Xu H**. A hybrid system for temporal information extraction from clinical text. J Am Med Inform Assoc. 2013, 20(5):828-35. [PMCID: PMC3756274]
22. McCoy AB, Wright A, Eysenbach G, Malin BA, Patterson ES, **Xu H**, Sittig DF. State of the art in clinical informatics: evidence and examples. *IMIA Yearbook of Medical Informatics*. 2013, 8(1):13-9. [PMID:23974543]
23. Mani S, Chen Y, Li X, Arlinghaus L, Chakravarthy AB, Abramson V, Bhave SR, Levy MA, **Xu H**, Yankeelov TE. Machine learning for predicting the response of breast cancer to neoadjuvant chemotherapy. *J Am Med Inform Assoc.* 2013, 20(4):688-95. [PMCID: PMC3721158]
24. Wei W, Cronin RM, **Xu H**, Lasko TA, Bastarache L, Denny JC. Development and evaluation of an ensemble resource linking medications to their indications. *J Am* Med Inform Assoc. 2013, 20(5):954-61. [PMCID: PMC3756263]
25. Tang B, Cao H, Wu Y, Jiang M, **Xu H**. Recognizing clinical entities in hospital discharge summaries using Structural Support Vector Machines with word representation features. *BMC Medical Informatics and Decision Making* 2013, 13(Suppl 1):S1. [PMCID: PMC3618243]
26. Chen Y, Cao H, Mei Q, Zheng K, **Xu H**. Applying Active Learning to Supervised Word Sense Disambiguation in MEDLINE. *J Am Med Inform Assoc.* 2013, 20(5):1001-6. [PMCID: PMC3756255]
27. Wiley LK, Shah A, **Xu H**, Bush WS. ICD-9 Tobacco Use Codes are Effective Identifiers of Smoking Status. *J Am Med Inform Assoc.* 2013, 20(4):652-8. [PMCID: PMC3721171]
28. Liu M, McPeek Hinz ER, Matheny ME, Denny JC, Schildcrout JS, Miller RA, **Xu H**. Comparative Analysis of Pharmacovigilance Methods in Detection of Adverse Drug Reactions from Electronic Medical Records. *J Am Med Inform Assoc.* 2013, 20(3):420-6. [PMCID: PMC3628053]
29. **Xu H**, Wu Y, Elhadad N, Stetson PD, Friedman C. A new clustering method for detecting rare sense of abbreviations in clinical notes. *J Biomed Inform.* 2012, 45(6):1075-83. [PMID: 22742938]
30. Liu M, Wu Y, Chen Y, Sun J, Zhao Z, Chen X, and **Xu H**. Large-scale prediction of adverse drug reaction by integrating chemical, biological, and phenotypic properties of drugs. *J Am Med Inform Assoc.* 2012. 19(e1): e28-e35. [PMCID:PMC3392844]
31. Denny JC, Schildcrout JS, Bowton EA, Gregg W, Pulley JM, Basford MA, Cowan J, **Xu H**, Ramirez AH, Crawford DC, Ritchie MD, Peterson JF, Masys DR, Wilke RA, Roden DM. Optimizing drug outcomes through pharmacogenetics: A case for preemptive genotyping. *Clin Pharmacol Ther.* 2012. 92(2):235-42. [PMID: 22739144]
32. Wu Y, Levy MA, Micheel CM, Yeh P, Tang B, Cantrell MJ, Cooreman SM, **Xu H**. Identifying the status of genetic lesions in cancer clinical trial documents using machine learning. *BMC Genomics.* 2012, 13 Suppl 8:S21 [PMCID: PMC3535695]
33. Han B, Chen XW, Talebizadeh Z, **Xu H**. Genetic studies of complex human diseases: characterizing SNP-disease associations using Bayesian networks. *BMC Syst Biol.* 2012. 6 Suppl 3: S14 [PMCID: PMC3524021]
34. Lu Y, **Xu H**, Peterson NB, Dai Q, Jiang M, Denny JC, Liu M. Extracting epidemiological exposure and outcome terms from literature using machine learning approaches. *Int J Data Min Bioinform*. 2012; 6(4):447-59. [PMID: 23155773]

35. Roden DM, **Xu H**, Denny JC, Wilke RA. Electronic Medical Records as a Tool in Clinical Pharmacology: Opportunities and Challenges. *Clin Pharmacol Ther.* 2012, Apr 25. [PMID:22534870]
36. Delaney JT, Ramirez AH, Bowton EA, Pulley JM, Basford MA, Schildcrout JS, Shi Y, Zink R, Oetjens M, **Xu H**, Cleator JH, Jahangir E, Ritchie MD, Masys DR, Roden DM, Crawford DC, Denny JC. Predicting clopidogrel response using DNA samples linked to an electronic health record. *Clin Pharmacol Ther.* 2012 Feb;91(2):257-63. [PMID: 22190063]
37. Birdwell KA, Grady B, Choi L, **Xu H**, Bian A, Denny JC, Jiang M, Vranic G, Basford M, Cowan JD, Richardson DM, Robinson MP, Ikizler TA, Ritchie MD, Stein CM, Haas DW. The use of a DNA biobank linked to electronic medical records to characterize pharmacogenomic predictors of tacrolimus dose requirement in kidney transplant recipients. *Pharmacogenet Genomics.* 2012 22(1):32-42. [PMCID: PMC3237759]
38. Ramirez AH, Shi Y, Schildcrout JS, Delaney JT, **Xu H**, Oetjens MT, Zuvich RL, Basford MA, Bowton EA, Jiang M, Speltz P, Zink R, Cowan J, Pulley JM, Ritchie MD, Masys DR, Roden DM, Crawford DC, Denny JC. Predicting warfarin dosage in European and African Americans using DNA samples linked to an electronic health record. *Pharmacogenomics*. 2012, 13(4):407-18. [PMCID: PMC3361510]
39. Sun J, **Xu H**, Zhao Z Network-assisted investigation of antipsychotic drugs and their targets. *Chem Biodivers*. 2012, 9(5): 900-10. [PMID:22589091]
40. Sun J, Wu Y, **Xu H**, Zhao Z. DTome: a web-based tool for drug-target interactome construction, *BMC Bioinformatics.* 2012, 13(Suppl 9): 57.
41. Carroll RJ, Thompson WK, Eyler AE, Mandelin AM, Cai T, Zink RM, Pacheco JA, Boomershine CS, Lasko TA, **Xu H**, Karlson EW, Perez RG, Gainer VS, Murphy SN, Ruderman EM, Pope RM, Plenge RM, Kho AN, Liao KP, Denny JC. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. *J Am Med Inform Assoc*. 2012, 19(e1):e162-9. [PMCID: PMC3392871]
42. Doan S, Collier N, **Xu H**, Pham HD, and Tu MP. Recognition of medication information from discharge summaries using ensembles of classifiers. *BMC Medical Informatics and Decision Making*. 2012, 12(1):36. [PMID: 22564405]
43. Chen Y, Mani S, **Xu H**. Applying active learning to assertion classification of concepts in clinical text. *J Biomed Inform* 2012, 45(2): 265-272. [PMCID: PMC3306548]
44. Wilke RA, **Xu H,** Denny JC, Roden DM, Krauss RM, McCarty CA, Davis RL, Skaar T, Lamba J, and Savova G. The emerging role of electronic medical records in pharmacogenomics. *Clin Pharmacol Ther.* 2011, 89(3): 379-86. [PMCID: PMC3204342]
45. Rosenbloom ST, Denny JC, **Xu H**, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *J Am Med Inform Assoc*. 2011, 18(2):181-6. [PMCID: PMC3116264]
46. Jiang M, Chen Y, Liu M, Rosenbloom ST, Mani S, Denny JC, **Xu H**. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. *J Am Med Inform Assoc*. 2011, 18(5):601-6. [PMCID: PMC3168315]
47. **Xu H**, Jiang M, Oetjens M, Bowton EA, Ramirez AH, Jeff JM, Basford MA, Pulley JM, Cowan JD, Wang X, Ritchie MD, Masys DR, Roden DM, Crawford DC, Denny JC. Facilitating pharmacogenetic studies using electronic health records and natural language processing: a case study of warfarin. *J Am Med Inform Assoc*. 2011; 18(4): 387-91. [PMCID: PMC3128409]

48. **Xu H**, AbdelRahman S, Lu Y, Denny JC, Doan S. Applying semantic-based probabilistic context free grammar to medical language processing – a preliminary study on parsing medication sentences. *J Biomed Inform* 2011, 44(6): 1068-75. [PMCID: PMC3226929]
49. **Xu H**, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx – A Medication Information Extraction System for Clinical Narratives. *J Am Med Inform Assoc*. 2010; 17(1):19-24. [PMCID: PMC2995636]
50. Denny JC, Peterson JF, Choma NN, **Xu H**, Miller RA, Bastarache L, Peterson NB. Development of a Natural Language Processing System to Identify Timing and Status of Colonoscopy Testing in Electronic Medical Records. *J Am Med Inform Assoc*. 2010; 17(4): 393-8. [PMCID: PMC2815478]
51. Doan S, Bastarache L, Klimkowski S, Denny JC, **Xu H**. Integrating Existing NLP Tools for Medication Extraction from Discharge Summaries. *J Am Med Inform Assoc.* 2010, 17:528-31. [PMCID: PMC2995674]
52. **Xu H,** Stetson P, Friedman C. Methods for Building Sense Inventories of Abbreviations in Clinical Notes. *J Am Med Inform Assoc*. 2009 16(1):103-108. [PMCID: PMC2605589]
53. Chen ES, Hripcsak G, **Xu H**, Markatou M, Friedman C. Automated Acquisition of Disease-Drug Knowledge from Biomedical and Clinical Documents. *J Am Med Inform Assoc*. 2008, 15(1):87-98. [PMCID: PMC2274872]
54. Tulipano KP, Tao Y, Millar WS, Zanzonico P, Kolbert K, **Xu H**, Yu H, Chen L, Lussier YA, Friedman C. Natural language processing and visualization in the molecular imaging domain. *J Biomed Inform.* 2007; 40:3, 270-281. [PMID: 17084109]
55. Fan JW, **Xu H,** Friedman C. Using Contextual and lexical features to restructure and validate the classification of biomedical concepts. *BMC Bioinformatics.* 2007; 8: 264. [PMCID: PMC2014782]
56. **Xu H**, Fan JW, Hripcsak G, Mendonça EA, Markatou M, Friedman C. Gene symbol disambiguation using knowledge-based profiles. *Bioinformatics*, 2007 23(8):1015-1022. [PMID: 17314123]
57. **Xu H**, Markatou M, Dimova R, Liu H, Friedman C. Machine learning and word sense disambiguation in the biomedical domain: design and evaluation issues. *BMC Bioinformatics*. 2006; 7:334. [PMCID: PMC1550263]
58. Lee HT, Krichevsky IE, **Xu H**, Ota-Setlik A, D'Agati VD, Emala CW. Local anesthetics worsen renal function after ischemia-reperfusion injury in rats. *Am J Physiol Renal Physiol.* 2004; 286(1):F111-9. [PMID: 14519592]
59. Lee HT, **Xu H**, Nasr SH, Schnermann J, Emala CW. A1 adenosine receptor knockout mice exhibit increased renal injury following ischemia and reperfusion. *Am J Physiol Renal Physiol.* 2004; 286(2):F298-306. [PMID: 14600029]
60. Lee HT, **Xu H**, Ota-Setlik A, Emala CW. Oxidant preconditioning protects human proximal tubular cells against lethal oxidant injury via p38 MAPK and heme oxygenase-1. *Am J Nephrol*. 2003; 23(5):324-33. [PMID: 12915776]
61. Lee HT, Ota-Setlik A, **Xu H**, D'Agati VD, Jacobson MA, Emala CW. A3 adenosine receptor knockout mice are protected against ischemia- and myoglobinuria-induced renal failure. *Am J Physiol Renal Physiol.* 2003; 284(2):F267-73. [PMID: 12388399]
62. Lee HT, **Xu H**, Siegel CD, Krichevsky IE. Local anesthetics induce human renal cell apoptosis. *Am J Nephrol.* 2003; 23(3):129-39. [PMID: 12586958]

Peer Reviewed Articles - Conference:

63. Chen Y, Wrenn J, **Xu H**, Spickard A, Habermann R, Powers J, Denny JC, Automated Assessment of Medical Students' Clinical Exposures according to AAMC Geriatric Competencies, *AMIA Annu Symp Proc* 2014 (Accepted).
64. Jiang M, Huang Y, Fan JW, Tang B, Denny J, **Xu H**. Parsing clinical text: how good are the state-of-the-art parsers? *ACM Eighth International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO)*, 2014, Shanghai, China. (* will be published in *BMC Medical Informatics and Decision Making* later).
65. Tao C, Sun J, Zheng W, Chen J, **Xu H**, Drug target prediction for colorectal cancer by combining ontology and network approaches, *International Conference on Biomedical Ontologies (ICBO)* 2014, accepted. (* will be published in *Database* later)
66. Chen G, Zhao J, Cohen T, Tao C, Sun J, **Xu H**, Bernstam E, Zeng J, Johnson A, Holla, Ann Balley V, Meric-Bernstam F and Zheng WJ, Standardized Pharmacological Class Profile in the Web Ontology Language (OWL), *International Conference on Biomedical Ontologies (ICBO)* 2014, accepted. (* will be published in *Database* later)
67. Madani S, Alemy R, Sittig DF, **Xu H**. Information extraction from clinical narratives: assessing ontology components. *International Conference on Biomedical Ontologies (ICBO)* 2014, accepted
68. Jiang M, Wu Y, Shah A, Priyanka P, Denny JC, **Xu H**. Extracting and standardizing medication information in clinical text – the MedEx-UIMA system. *AMIA Clinical Research Informatics Summit*, 2014.
69. Zeng J, Wu Y, Balley, A, Johnson A, Halla V, Bernstam EV, **Xu H**, Meric-Bernstam F. Adapting a natural language processing tool to facilityate clinical trial curation for personalized cancer therapy. *AMIA Translational Bioinformatics Summit*, 2014.
70. Tang B, Wang X, Wu Y, Jiang M, Wang J, **Xu H**. Recognizing Chemical Entities in Biomedical Literature using Conditional Random Fields and Structured Support Vector Machines. *BioCreative Challenge Evaluation Workshop* 2013 vol. 2, 70
71. Wu Y, Tang B, Jiang M, Moon S, Denny JC, **Xu H**. Clinical Acronym/Abbreviation Normalization using a Hybrid Approach. 2013. Proceedings of CLEF 2013 Evaluation Labs and Workshop.
72. Wu Y, Denny JC, Rosenbloom ST, Miller RA, Giuse D, Song M, **Xu H.** A Prototype Application for Real-time Recognition and Disambiguation of Clinical Abbreviations. *ACM Seventh International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO)*, 2013, San Francisco, USA.
73. Tang B, Wu Y, Jiang M, and **Xu H**. A Machine Learning based System for Disorder Concept Extraction. 2013. Proceedings of CLEF 2013 Evaluation Labs and Workshop.
74. Moon S, Berster BT, **Xu H**, Cohen T. Word sense disambiguation of clinical abbreviations with hyperdimensional computing. *AMIA Annu Symp Proc.* 2013. 16;1007-16. [PMCID: PMC3900125]
75. Wu Y, Lei J, Wei W, Tang B, Denny JC, Rosenbloom ST, Miller RA, Giuse DA, Zhang K, **Xu H**. Analyzing differences between Chinese and English clinical text: a cross-institution comparison of discharge summaries in two languages. *Stud Health Technol Inform*, 2013, 192:662-6. [PMID: 23920639]
76. Sun S., Zhou X., Denny JC, Rosenbloom CT, **Xu H**. Messaging to Your Doctors: Understanding Patient-Provider Communications via a Portal System. *Proceedings of ACM Conference on Human Factors in Computing Systems (SIG'CHI),* 2013, Paris, France. 1739-48.
77. Tang B, Can H, Wu Y, Jiang M, **Xu H**. Clinical Entity Recognition using Structural Support Vector Machines with Rich Features. *ACM Sixth International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO)*, 2012, 13-20.

78. **Xu H**, Stetson PD, Friedman C. Combining corpus-derived sense profiles with estimated frequency information to disambiguate clinical abbreviations. *AMIA Annu Symp Proc.* 2012. 1004-13. [PMCID PMC3540457]
79. Wu Y, Denny JC, Rosenbloom ST, Miller RA, Giuse DA, **Xu H**. A comparative study of current clinical natural language processing systems on handling abbreviations in discharge summaries. *AMIA Annu Symp Proc.* 2012. 997-1003.
80. Liu M, Shah A, Min J, Peterson NB, Dai Q, Aldrich MC, Chen Q, Bowton EA, Liu H, Denny JC, **Xu H**. A study of transportability of an existing smoking status detection module across institutions. *AMIA Annu Symp Proc.* 2012. 577-86.
81. Jiang M, Denny JC, Tang B, Cao H, **Xu H**. Extracting semantic lexicons from discharge summaries using machine learning and c-value method. *AMIA Annu Symp Proc.* 2012. 409-16.
82. Wu Y, Liu M, Zheng W, Zhao Z, **Xu H**. Ranking gene-drug relationships in biomedical literature using latent dirichlet allocation. *Pac Symp Biocomput.* 2012: 422-33. [PMID: 22174297]
83. **Xu H**, Fu Z, Shah A, Chen Y, Peterson NB, Chen Q, Mani S, Levy MA, Dai Q,Denny JC. Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases. *AMIA Annu Symp* Proc. 2011, 1564-72. [PMCID: PMC3244156]
84. Liu M, Kawai VK, Stein CM, Denny JC, Roden DM, **Xu H**. Modeling drug exposure data in electronic medical records: an application to warfarin. *AMIA Annu Symp Proc.* 2011, 815-23. [PMCID: PMC3243123]
85. Wu Y, Rosenbloom ST, Denny JC, Miller RA, Mani S, Giuse DA, **Xu H**. Detecting abbreviations in discharge summaries using machine learning methods. *AMIA Annu Symp Proc.* 2011, 1541-9. [PMCID: PMC3243185]
86. **Xu H**, AbdelRahman S, Jiang M, Fan JW, Huang Y. An Initial Study of Full Parsing of Clinical Text using the Stanford Parser. *International Workshop on Biomedical and Health Informatics, IEEE Conference of Bioinformatics and Biomedicine (BIBM)*, 2011.
87. **Xu H,** Doan S, Birdwell KA, Cowan JD, Vincz AJ, Haas DW, Basford MA, Denny JC. An automated approach to calculating the daily dose of tacrolimus in electronic health records. *AMIA Summits Transl Sci Proc.* 2010:71-5. [PMCID: PMC3041548]
88. Denny JC, Speltz P, Maddox R, Stein G, **Xu H**, Spickard A. Comparing Content Coverage in Medical Curriculum to Trainee-Authored Clinical Notes. *AMIA Annu Symp Proc.* 2010, 157-161. [PMCID: PMC3041398]
89. Doan S and **Xu H**. Recognizing Medication related Entities in Hospital Discharge Summaries using Support Vector Machine. *COLING 2010*, the 23rd International Conference on Computational Linguistics, 259-266.
90. **Xu H**, Lu Y, Jiang M, Liu M, Denny JC, Dai Q, Peterson NB. Mining Biomedical Literature for Terms related to Epidemiologic Exposures. *AMIA Annu Symp Proc.* 2010, 897-901. [PMCID: PMC3041399]
91. Fan JW, **Xu H**, Friedman C. Using Distributional Analysis to Semantically Classify UMLS Concepts. *In Proceedings of Medinfo.* 2007; 519-23. [PMID: 17911771]
92. **Xu H,** Fan JW, Friedman C. Combine multiple evidence for gene symbol disambiguation. *ACL 2007, BioNLP Workshop*, p41-48.
93. **Xu H,** Stetson P, Friedman C. A Study of Abbreviations in Clinical Notes. *AMIA Annu Symp Proc.* 2007; 821-5.
94. **Xu, H**, Anderson, K, Grann, V, Friedman, C. Facilitating Cancer Research using Natural Language Processing of Pathology Reports. *In Proceedings of Medinfo*. 2004; 565-72. [PMID: 15360876]

Editorials and Invited Articles:

95. Liu M, Matheny ME, Hu Y, **Xu H**. Data mining methodologies for pharmacovigilance. *ACM SIGKDD Explorations Newsletter*. 2012; 14(1), 35-42.
96. Huang Y, Zhao Z, **Xu H**, Shyr Y, Zhang B. Advances in systems biology: Computational algorithms and applications. *BMC Systems Biology*, 2012, 6 (Suppl S3):S1
97. Zhao Z, Zhang B, Shyr Y, Huang Y, **Xu H**. Genomics in 2012: challenges and opportunities in the next generation sequencing era. *BMC Genomics,* 2012, 13 (Suppl S8):S1
98. Zhao Z, **Xu H**, Zhang B, Huang Y. Advances in intelligent biology and medicine. *International Journal of Computational Biology and Drug Design (IJCBDD)*, 2013, 6(1-2):1-4
99. Zhang B, Huang Y, McDermott JE, Posey RH, **Xu H**, Zhao Z. Interdisciplinary dialogue for education, collaboration, and innovation: Intelligent biology and medicine in and beyond 2013. *BMC Genomics*, 2013, 14 (Suppl S8):S1
100. Zhao Z, Zhang B, Huang Y, **Xu H**, McDermott JE. Computational methods for biomedical big data. *International Journal of Computational Biology and Drug Design (IJCBDD)*. 2014, 7(2-3):97-101
101. McDermott JE, Huang Y, Zhang B, **Xu H**, Zhao Z. Integrative genomics and computational systems medicine. *BioMed Research International.* 2014, Article ID: 945253

Book Chapters:

1. Liu M, Hu Y, Matheny ME, and **Xu H**. Artificial Intelligence Approaches for Drug Safety Surveillance and Analysis. *Medical Applications of Artificial Intelligence,* Taylor & Francis Group, 2013.
2. Denny JC and **Xu H**. Linking genomic and clinical data for discovery and personalized care. *Methods in Biomedical Informatics: A Pragmatic Approach*. Academic Press, 2013
3. **Xu H** and Denny JC. Unlocking Information in Electronic Health Records Using Natural Language Processing: A Case Study in Medication Information Extraction. *Text Mining of Web-based Medical Content,* De Gruyter, 2014

Abstracts/Posters:

1. Chen Y, Zhang Y, Mei Q, Lasko T, Denny JC, **Xu H.** A preliminary study of coupling transfer learning with active learning for clinical named entity recognition between two institutions. *AMIA Annu Symp Proc*, 2014, Washington DC, US.
2. Moon S, Cohen T, **Xu H**. A study of synonym extraction from clinical texts using sematic vector models. *AMIA Annu Symp Proc*, 2014, Washington DC, US.
3. Sun J, Cui T, Zhu KW, Zheng J, Chen J, **Xu H**. an integrative framework for drug target prediction and repurposing. *AMIA Annu Symp Proc*, 2014, Washington DC, US.
4. Wu Y, Wright A, **Xu H**, McCoy AB, Sittig DF. Development of a unified computable problem-medication knowledge base. *AMIA Annu Symp Proc*, 2014, Washington DC, US.
5. Jiang M, Huang Y, Fan JW, Yang EW, **Xu H**. Building a Treebank of hospital discharge summaries. *AMIA Annu Symp Proc*, 2014, Washington DC, US.

6. **Xu H**, Aldrich MC, Chen Q, Peterson NB, Dai Q, Levy M, Shah A, Xue H, Li Y, Friedman C, Liu H, Denny JC. Electronic health record data suggests metformin improves cancer survival: A new model for drug repurposing studies. *AMIA Annu Symp Proc*, 2012, Chicago, IL.
7. Sun S, Zhou X, Denny JC, Rosenbloom CT, **Xu H**. (2012). Understanding Patient-Provider Communication Entered Via a Patient Portal System. *ASIS&T* 2012, Baltimore, Maryland.
8. Sun S, Zhou X, Denny JC, Rosenbloom CT, **Xu H**. Understanding Patient-Provider Communication via a Patient Portal. *AMIA Annu Symp Proc*, 2012, Chicago, IL.
9. Tang B, Wu Y, Jiang M, **Xu H**. Clinical Entity Recognition using Structural Support Vector Machines. *AMIA Annu Symp Proc*, 2012, Chicago, IL.
10. Shah A, Jiang M, Wu Y, Denny JC, **Xu H**. MedEx-UIMA – An Open-Source System for Medication Information Extraction from Clinical Text. *AMIA Annu Symp Proc*, 2012, Chicago, IL.
11. Ramirez AH, Shi Y, Schildcrout, Delaney JT, **Xu H**, Oetjens MT, Zuvich RL, Basford MA, Bowton E, Jiang M, Zink R, Cowan J, Pulley JM, Ritchie MD, Peterson JF, Masys DM, Roden DM, Crawford DC, Denny JC. Predicting warfarin dosage in European and African Americans using DNA samples linked to an electronic health record. *International Congress of Human Genetics*, 2011.
12. Liu M, McPeek Hinz ER, Denny JC, Mani S, Chen Y, Schildcrout JS, Miller RA, **Xu H**. A Feasibility Study in Detection of Adverse Drug Effects using Inpatient Laboratory Test Results. *International Biomedical Informatics Summit at Peking University*, Beijing, China, 2011
13. Birdwell KA, Grady B, Choi L, **Xu H**, Denny JC, Jiang M, Ritchie MD, Vranic G, Cowan JD, Basford BA, Stein CM, Ikizler TA, Haas DW. Use of a Novel Genomics/Informatics Resource to Identify Predictors of Tacrolimus Blood Concentrations in Kidney Transplant Recipients. *American Transplant Congress*, Philadelphia, PA, 2011.
14. Schildcrout JS, Denny JC, Bowton E, Gregg W, Pulley JM, Basford MA, Cowan J, **Xu H**, Ramirez AH, Crawford DC, Ritchie MD, Peterson JF, Masys DR, Wilke RA, Roden DM. Optimizing drug outcomes through pharmacogenetics: A case for preemptive genotyping. *International Congress of Human Genetics*, 2011.
15. Denny JC, Bastarache L, Ritchie MD, Basford MA, Pulley JM, Zuvich RL, Peissig PL, Carrell D, Pathak P, Rasmussen L, Pacheco J, Kho A, Weston N, Pendergrass S, **Xu H**, Li R, Manolio TA, Kullo IJ, Chute CG, Chisolm RL, Larson EB, McCarty CA, Masys DR, Roden DM, Crawford DC. A phenome-wide analysis of SNPs in the National Human Genome Research Institute genome-wide association catalog. *International Congress of Human Genetics*, 2011.
16. **Xu H**, Cunningham AJ, Roden DM, Stein CM. Facilitating Pharmacogenetic studies of warfarin outcomes using informatics methods. *ACRT* 2010
17. Liu M, Denny JC, Mani S, Chen Y, Hu Y, **Xu H**. Identifying potential drugs that induce QT prolongation using electronic medical records. *UT-ORNL-KBRIN Bioinformatics Summit* 2010.
18. Jiang M and **Xu H**. Fine-grained semantic analysis of numbers in clinical text. *AMIA Annu Symp Proc*, 2010.
19. Ramirez AH, **Xu H**, Oetjens M, Jeff JM, Zuvich R, Basford MA, Bowton E, Pulley JM, Ritchie MD, Crawford DC, Masys DR, Roden DM, Denny JC. Identifying genotype-phenotype relations in electronic medical record systems: application to warfarin pharmacogenomics. *American Heart Association*, 2010.

20. **Xu H**, Krupke D, Blake J, Friedman C. A Natural Language Processing (NLP) Tool to Assist in the Curation of the Laboratory Mouse Tumor Biology Database. *AMIA Annu Symp Proc*. 2006:1150.
21. **Xu H**, Friedman C. Facilitating Research in Pathology using Natural Language Processing. *AMIA Annu Symp Proc*. 2003:1057.
22. Lee HT, **Xu H,** Emala CW. Sevoflurane Protects Against Renal Ischemic Reperfusion Injury in Rats. *ASA Meeting* 2002.